



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJIRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Issue 3, March 2023

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.423

9940 572 462

6381 907 438

ijirset@gmail.com

www.ijirset.com

A Survey on Sales Prediction Using Machine Learning

Rohit Athawale, Kishor Deshmukh, Prathamesh Dasade, Aniket Nagul, Prof.S.K.Shinde

Student, Department of Computer Engineering, K J College of Engineering and Management Research, Pune, India

Student, Department of Computer Engineering, K J College of Engineering and Management Research, Pune, India

Student, Department of Computer Engineering, K J College of Engineering and Management Research, Pune, India

Student, Department of Computer Engineering, K J College of Engineering and Management Research, Pune, India

Associate Professor, Department of Computer Engineering, K J College of Engineering and Management Research, Pune, India

ABSTRACT: This paper discusses the use of machine learning techniques for predicting sales in the retail industry, which is essential for making informed decisions about inventory management, marketing strategies, and overall business growth. Various machine learning algorithms, such as XGBoost, linear regression, neural networks, and CatBoost, have been used in previous studies to forecast sales. Features such as product type, price, promotion, and seasonality were used to train the models. The study also discusses the practical implications of these models for inventory management and sales forecasting in the retail industry and machine learning models can accurately predict future sales by learning from historical sales data.

I. INTRODUCTION

Sales prediction is a crucial aspect of running a successful business. Accurately forecasting future sales enables businesses to make informed decisions about inventory management, marketing strategies, and overall business growth. With the advancements in technology, machine learning has emerged as a powerful tool for sales prediction. Machine learning models can be trained on historical sales data to learn patterns and trends and make accurate predictions about future sales.

The main goal of sales prediction using machine learning is to develop a predictive model that can forecast future sales with a high degree of accuracy. This involves collecting and pre-processing data, selecting relevant features, selecting and training a machine learning model, evaluating its performance, and deploying it in production. The model can then be integrated with other systems or applications to automate sales prediction.

II. RELATED WORKS

In the paper [1], the authors developed a sales forecasting model using the XGBoost algorithm. The XGBoost algorithm is a popular machine learning algorithm that is known for its high accuracy and scalability. The authors collected sales data from an online retailer over a period of two years and preprocessed the data to prepare it for analysis. They then selected relevant features, such as time, price, and product category, and used them to train the XGBoost model. The authors evaluated the performance of the model using various metrics, such as mean absolute error (MAE), mean squared error (MSE), and root mean squared error (RMSE). They compared the performance of the XGBoost model with that of other machine learning models, such as linear regression and random forest, and found that the XGBoost model outperformed them in terms of accuracy.

The paper [2] presents a method for predicting the sales value in online shopping using linear regression. The authors collected data from an online shopping website and used it to train a linear regression model. They used various features such as the price of the product, the product category, the time of day, the day of the week, and the number of clicks on the product to predict the sales value. The authors used the R programming language and various libraries such as ggplot2, caret, and e1071 to implement their model. They evaluated their model using various performance metrics such as the mean absolute error, mean squared error, and R-squared value. The results showed that the linear regression model was able to predict the sales value with reasonable accuracy.



The paper [3] discusses a method for predicting sales using online sentiment analysis and regression modelling. The authors propose a framework that combines sentiment analysis of online reviews and ratings with regression modelling techniques to predict future sales of a product or service. The authors explain that this approach can be useful for businesses to better understand customer feedback and make informed decisions about product development, marketing, and sales strategies. The authors tested their framework on a real-world dataset of hotel reviews collected from the TripAdvisor website. They used sentiment analysis techniques to classify the reviews as positive, negative, or neutral and then applied regression modelling techniques to predict hotel occupancy rates based on the sentiment scores. The authors claim that their approach outperformed traditional methods of sales prediction.

In this paper [4], they proposed Neural Network model to predict Walmart sales on Walmart sales dataset provided by Kaggle website: M5 Forecasting - Accuracy, Estimate the unit sales of Walmart retail goods. They used TensorFlow and Keras to train our model and experiments show our NN model performs better than Linear Regression algorithm and SVM algorithm. Additionally, they utilize SHAP to analyze the feature importance of our NN model.

In the paper [5], author proposed new method to predict Walmart sales. They combine XGBoost algorithm and feature engineering processing on Walmart sales dataset available on Kaggle platform. Experiments show their method can effectively mine the features of different dimensions and performs better than logistic regression algorithm and ridge algorithm.

The paper [6] discusses the use of CatBoost, a machine learning algorithm, for sales forecasting. Sales forecasting is an important task in business management, as it allows companies to predict future sales and make better decisions about inventory, marketing, and other factors. The authors used a dataset from an e-commerce company to train and test the CatBoost model. They used various features such as the product category, brand, price, and time of purchase to predict sales. The performance of the model was evaluated using metrics such as the mean absolute error (MAE) and the root mean squared error (RMSE). The results of the study showed that CatBoost outperformed other machine learning algorithms such as Random Forest and Gradient Boosting in terms of prediction accuracy. The authors also discussed the importance of feature engineering and hyperparameter tuning in improving the performance of the model.

The paper [7] The authors used a dataset from a retail business to train and test the machine learning models. They used various features such as product category, price, time of purchase, and customer demographics to predict sales and revenue. The performance of the models was evaluated using metrics such as mean absolute percentage error (MAPE), mean squared error (MSE), and root mean squared error (RMSE). The study compares the performance of various machine learning algorithms such as Linear Regression, Random Forest, and XGBoost. The results showed that XGBoost outperformed other algorithms in terms of prediction accuracy.

III. METHODOLOGY

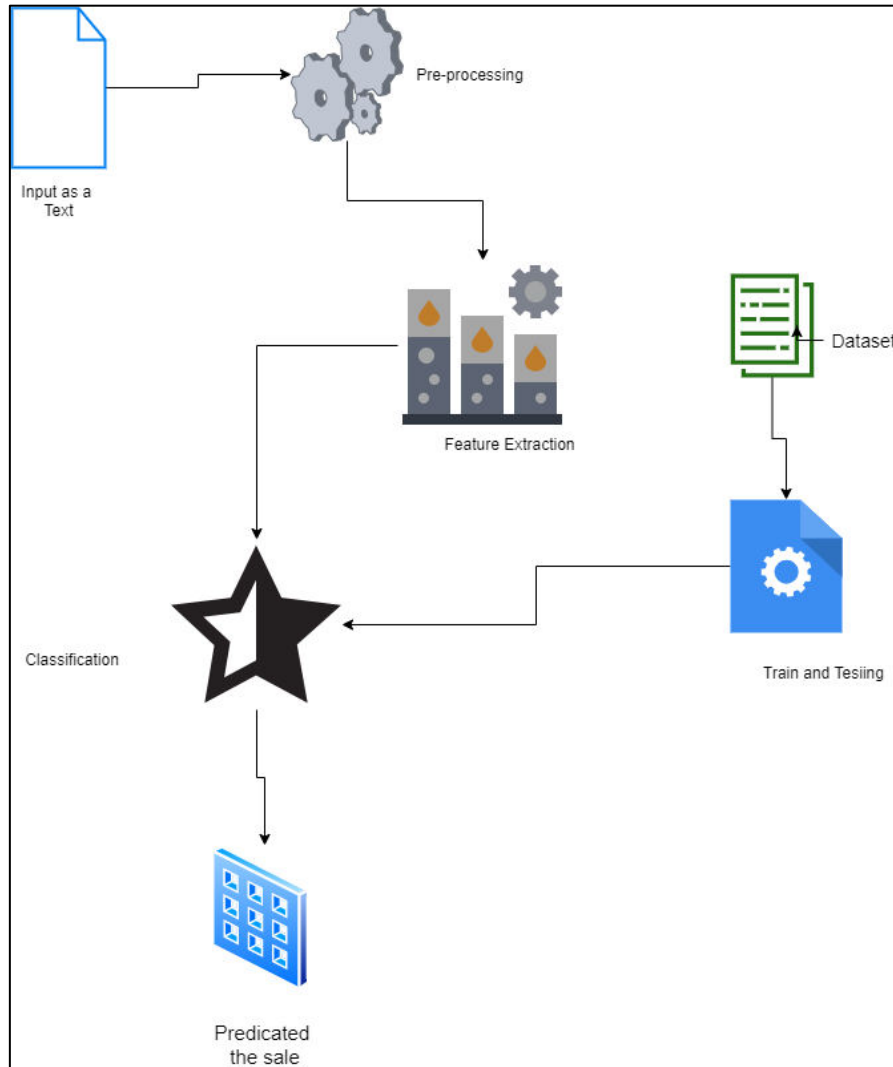


Figure 1: Block Diagram of System

IV. PROPOSED SYSTEM

Sales prediction is an important task for businesses, as it helps them make informed decisions about inventory management, marketing strategies, and overall business growth. Machine learning can be used to build predictive models that can help businesses forecast their sales accurately. Here is a proposed system for sales prediction using machine learning:

Data Collection: The first step is to collect data on historical sales. This data can be obtained from internal systems or external sources. It is important to collect data over a period of time to identify patterns and trends in sales.

Data Pre-processing: The next step is to pre-process the data. This involves cleaning the data, removing duplicates, handling missing values, and transforming the data into a format suitable for analysis.

Feature Selection: Once the data is pre-processed, the next step is to select relevant features. This involves identifying the features that have a significant impact on sales. For example, features such as price, promotion, seasonality, and customer demographics may be relevant.



Model Selection: The next step is to select a machine learning model that is suitable for sales prediction. There are various models that can be used for sales prediction, such as linear regression, decision trees, random forests, and neural networks.

Model Training: Once the model is selected, the next step is to train the model using historical sales data. The model is trained to learn the patterns and trends in the data and to make accurate predictions.

Model Evaluation: After the model is trained, it is important to evaluate its performance. This involves testing the model on a validation set and measuring its accuracy, precision, recall, and F1 score.

Model Deployment: Once the model is evaluated and its performance is satisfactory, it can be deployed in production. The model can be integrated with other systems or applications to automate sales prediction.

V. CONCLUSION

There are several benefits of using machine learning for sales prediction. Firstly, it enables businesses to make informed decisions based on data-driven insights. It can also help businesses identify areas of improvement and optimize their marketing strategies. Then it improves overall business efficiency by streamlining inventory management and supply chain operations.

Sales prediction using machine learning is an effective way for businesses to forecast future sales and make informed decisions. By leveraging historical sales data and advanced machine learning algorithms, businesses can gain valuable insights into their sales patterns and make data-driven decisions to drive business growth.

REFERENCES

- [1] X. dairu and Z. Shilong, "Machine Learning Model for Sales Forecasting by Using XGBoost," 2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE), Guangzhou, China, 2021, pp. 480-483, doi: 10.1109/ICCECE51280.2021.9342304.
- [2] G. T., R. Choudhary and S. Prasad, "Prediction of Sales Value in Online shopping using Linear Regression," 2018 4th International Conference on Computing Communication and Automation (ICCCA), Greater Noida, India, 2018, pp. 1-6, doi: 10.1109/CCAA.2018.8777620.
- [3] S. K. Punjabi, V. Shetty, S. Pranav and A. Yadav, "Sales Prediction using Online Sentiment with Regression Model," 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2020, pp. 209-212, doi: 10.1109/ICICCS48265.2020.9120936.
- [4] J. Chen, W. Koju, S. Xu and Z. Liu, "Sales Forecasting Using Deep Neural Network And SHAP techniques," 2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE), Nanchang, China, 2021, pp. 135-138, doi: 10.1109/ICBAIE52039.2021.9389930.
- [5] Y. Niu, "Walmart Sales Forecasting using XGBoost algorithm and Feature engineering," 2020 International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE), Bangkok, Thailand, 2020, pp. 458-461, doi: 10.1109/ICBASE51474.2020.00103.
- [6] J. Ding, Z. Chen, L. Xiaolong and B. Lai, "Sales Forecasting Based on CatBoost," 2020 2nd International Conference on Information Technology and Computer Application (ITCA), Guangzhou, China, 2020, pp. 636-639, doi: 10.1109/ITCA52113.2020.00138.
- [7] M. A. R. Mia, M. A. Yousuf and R. Ghosh, "Business Forecasting System using Machine Learning Approach," 2021 2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST), DHAKA, Bangladesh, 2021, pp. 314-318, doi: 10.1109/ICREST51555.2021.9331114.



INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.118



ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 **9940 572 462**  **6381 907 438**  **ijirset@gmail.com**



www.ijirset.com

Scan to save the contact details